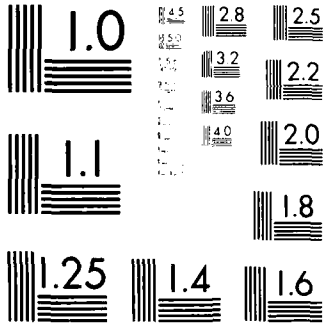
AD-A170 003 A RATIONAL APPROACH TO THE OPTIMAL DESIGN OF DRUGS(U) 1/1
GEORGIA UNIV ATHENS DEPT OF CHEMISTRY M RANDIC ET AL.
16 JUL 86 TR-23 N00014-85-K-0365

UNCLASSIFIED F/G 6/15 NL



(12)

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report No. 23	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A RATIONAL APPROACH TO THE OPTIMAL DESIGN OF DRUGS		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) M. Randić, B. Jerman-Blazić, S.C. Grossman, and D.H. Rouvray		8. CONTRACT OR GRANT NUMBER(s) N00014-84-K-0365
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Georgia Department of Chemistry Athens, GA 30602		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR051-861
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Department of the Navy Arlington, VA 22217		12. REPORT DATE 7/16/86 13. NUMBER OF PAGES 12
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) 15a. DECLASSIFICATION/ DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

To be published in Proceedings of the Fifth International Conference on 'Mathematical Modeling in Science and Technology'

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Drug design, graph theory, optimization techniques.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Recent advances in the rational design of drug molecules based on a graph-theoretical approach are briefly reviewed. Graph theory has not been widely recognized to date as an effective alternative to the empirical procedures currently prevailing in the development of new drugs. Moreover, the problems confronting researchers in this field are daunting in their great complexity. We advocate here a novel yet simple mathematical formalism which opens up a promising new avenue of research. After outlining the fundamental premises of our method, we exemplify it by discussing the

(continued)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

#20.

characterization, comparison, and quantification of similarity among individual molecules. It is indicated how the essential bioactive component in molecules of compounds displaying similar pharmacological behavior may be identified. We conclude by describing the pharmacological classification of 18 compounds, all of which are structurally similar but which exhibit several differing types of bioactivity.

OFFICE OF NAVAL RESEARCH

Contract N00014-84-K-0365

TECHNICAL REPORT NO. 23

A Rational Approach to the Optimal Design of Drugs

by

M. Randić, B. Jerman-Blazić, S.C. Grossman, and D.H. Rouvray

Prepared for publication in

Proceedings of the Fifth International Conference on

'Mathmatical Modeling in Science and Technology'

University of Georgia
Department of Chemistry
Athens, Georgia 30602

July 16, 1986

A RATIONAL APPROACH TO THE OPTIMAL DESIGN OF DRUGS

M. Randić[a], B. Jerman-Blazič[b], S.C. Grossman[c], and D.H. Rouvray[d*]

[a]Department of Mathematics and Computer Science, Drake University, Des Moines, Iowa 50311, and Ames Laboratory - DOE, Iowa State University, Ames, Iowa 50011, U.S.A.

[b]Jozef Stefan Institute, P.O.B. 199, 61001 Ljubljana, Slovenia, Yugoslavia.

[c]Department of Chemistry and Computer Science, North High School, 4323 North 37th Street, Omaha, Nebraska 68111, U.S.A.

[d]Department of Chemistry, University of Georgia, Athens, Georgia 30602, U.S.A.

Abstract. Recent advances in the rational design of drug molecules based on a graph-theoretical approach are briefly reviewed. Graph theory has not been widely recognized to date as an effective alternative to the empirical procedures currently prevailing in the development of new drugs. Moreover, the problems confronting researchers in this field are daunting in their great complexity. We advocate here a novel yet simple mathematical formalism which opens up a promising new avenue of research. After outlining the fundamental premises of our method, we exemplify it by discussing the characterization, comparison, and quantification of similarity among individual molecules. It is indicated how the essential bioactive component in molecules of compounds displaying similar pharmacological behavior may be identified. We conclude by describing the pharmacological classification of 18 compounds, all of which are structurally similar but which exhibit several differing types of bioactivity.

Keywords. Drug design; graph theory; optimization techniques.

INTRODUCTION

The history of drug development abounds in examples of major discoveries being made either serendipitously or as a result of following totally erroneous procedures (Burger, 1983). In spite of this circumstance, however, it was recognized very early on that the bioactivity of drug molecules was dependent upon the presence of special structural features in such molecules. It was pointed out by Crum Brown and Fraser (1868), for instance, that the quaternary ammonium group was essential for the blocking activity of curare-type drugs. Exactly one hundred years ago, Paul Ehrlich (1885), the founder of modern medicinal chemistry, elucidated the role played by enzymes in living systems. He thereby simplified the problem of drug interaction from one involving the study of cellular complexity to one involving complexity at no more than the molecular level. Ehrlich's work laid the foundation for our current theories of drug action, drug metabolism, and drug resistance.

Unfortunately, even today, very little is known about the mechanism of drug action or the underlying dynamics. The principal reason for this lacuna is a lack of knowledge of the structure of the relevant enzymes and an absence of detailed descriptions of their active sites. By contrast, most drugs can be viewed as small molecules (with a few notable exceptions) and are thus much more accessible to study. The fundamental problem in bioactivity studies therefore resolves itself into one of investigating the interaction of a relatively small, well characterized molecule with an unknown large protein molecule. Clearly, this represents an exceedingly difficult problem given the current level of our knowledge. Moreover, until our understanding of the participating protein and enzyme molecules approaches that existing for small molecules, and until our knowledge of all the intermediate steps which occur in an organism after administration of the drug becomes very detailed, such difficulties are likely to remain with us. This observation necessarily implies that we are very far from a situation where the use of any kind of rigorous theoretical technique could be comtemplated.

In coming to terms with the present state of affairs, pragmatism would seem to point in the direction of sacrificing our curiosity on how the whole process evolves

and focusing instead on what is now within our reach. If we adopt a typical systems analysis approach (White and Tauber, 1969), we may probe the system by means of an appropriate input (drug) and then examine the resultant output (pharmacological activity). Approximate schemes, empirical rules, statistical methods, and mathematical modelling might thus appear as the only reasonable routes to encompassing the vast amount of data on drugs which have accumulated over the years. The situation confronting us, however, is perhaps not as bleak as it may at first seem, for there is a good deal of evidence which indicates that apparently similar compounds exhibit closely similar pharmacological activities.

One of the first to recognize the relationship between the structure and activity of drugs was Emil Fischer (1894) in a paper entitled "Influence of configuration on the action of enzymes." The basic model that he put forward assumed that enzymes have recognition sites, i.e. receptor locations, that are highly specific structurally. Binding to a host, i.e. a drug, would be possible only if essential structural fragments in the drug molecule match up precisely with those at the receptor site. In more informal terms, this matching can be described in terms of a 'lock and key' analogy, with the drug playing the role of the key. The current status of medicinal chemistry can be summed up by stating that the available 'keys' are being employed to probe unknown 'locks' with a view to constructing improved 'keys' that will better fit the 'locks'.

In mathematical parlance, the above represents an example of a reconstruction problem: by collecting a fair number of responses, one tries to determine the optimal input. By inversion of its own connectivity, an optimal key molecule would certainly be able to provide valuable information about the structure of its receptor. In practice, once a reliable lead compound, i.e. a structure that triggers a useful response, has been identified, the next problem of selecting structures with enhanced biological activity would not be soluble without some guidelines as to the method of picking out the small number of highly active molecules from the usually enormous number of possible candidates. There is an astronomical number of combinatorial possibilities associated with even a modest number of substitution sites on a molecule and (say) a dozen or more potential substituents. Thus, starting from a given lead molecule, the essential task becomes one of devising some scheme whereby those few candidate structures which can function even more effectively as drugs than the lead can be recognized.

THE FUNDAMENTAL POSTULATE

Currently, two fundamentally different philosophies underlie the various approaches to the rational design of drugs. The first involves considering a large data set of compounds and reducing the size of the set by means of a number of empirical schemes, all of which are based essentially on statistical analysis. This reduces the problem to one of lower dimension and normally gives an indication of which parameters are critical. Once established, these parameters can be employed in the prediction of novel candidate drug molecules. Representative methods based on this school of thought include pattern recognition (Stuper et al., 1979) and regression analysis (Hansch, 1969). The second philosophy entails considering a small data set of compounds, the aim now being recognition of the degree of similarity between compounds of similar pharmacologic or therapeutic value. Exclusive use is made here of structural parameters for the description of the drug molecules, with the emphasis falling on the mathematical properties of the structures involved. Comparison of structures having similar mathematical properties is undertaken on the assumption that such structures will also display similar physical, chemical, and biological properties.

The fundamental basis of the second school of thought may be expressed in terms of the following postulate (Randić, 1985a):

POSTULATE: Structures which display substantial similarity in their mathematical properties will also display considerable similarity in their physical, chemical, and biological properties.

This postulate has a number of very important implications, each of which we now outline and assess in some detail:

(i) The various natural properties of chemical species may be characterized in purely mathematical terms. This is wellknown to be the case, witness the widespread use of topological indices (Bonchev, 1983) in the description of many natural phenomena. It was first suggested by Rouvray (1973) that topological indices might be used as mathematical descriptors for candidate molecules in drug design studies. The feasibility of this type of approach has been amply demonstrated in recent years (Kier and Hall, 1976);

(ii) The natural properties of chemical species are merely reflections of inherent mathematical properties of the structures concerned. According to this view, chemical species characterized by similar mathematical descriptors will be

possessed of similar physical, chemical, and biological properties. This statement is equivalent to saying that, if the mathematical descriptors of two structures are closely similar, the structures concerned will behave as isoteres, i.e. molecules which have related physicochemical properties and which exhibit broadly similar bioactivity (Langmuir, 1919; Thornber, 1979). Such a formulation places the focus of interest on the mathematical properties of structures and indicates that the natural properties may be compared and predicted by assessing the mathematical features of the structures concerned;

(iii) In going from structure to structure among molecules which are closely similar, it is postulated that there exist a rough continuum in the physical, chemical, and biological properties of the various species. Although it can never be strictly accurate to refer to a continuum when reference is made to discrete objects, i.e. molecules, it is our contention that very small changes in the neighborhood relations within the set of molecules will result in only minor changes in their natural properties. Thus, provided an appropriate set of structures is chosen, a more or less continuous range of properties can be generated without any significant gaps and with no abrupt changes;

(iv) Based on the above statement, it follows that, by making a suitable selection of structures and their substituents, any desired range of natural properties can be realized for a specific set of structures. In the light of the foregoing, the initial selection process would have to entail characterization of the mathematical nature of the candidate structures and the collection together of those structures which display only slight differences in terms of their mathematical descriptors. The use of mathematical criteria for the purpose of clustering structures around some defined natural property will form the subject matter of most of the rest of this presentation.

Although our postulate serves as an effective paradigm in specifying how chemical structures may be clustered around some natural property of interest, it does not reveal how the mathematical characteristics will be manifested in terms of the various natural properties of the set of chemical compounds investigated. In this respect, it is quite unlike the axioms of quantum theory or the laws of classical physics. Moreover, it should be mentioned that the postulate does not necessarily apply to individual compounds which are regarded as forming a continuum. In appropriate contexts, the postulate can refer to composite systems such as a drug-enzyme pair. A major criticism to our type of approach has been that chiral structures have the same mathematical properties (since they are identical in all respects apart from irrelevant mirror reflections) yet display dramatically different biological activities. Such criticism is invalid because the biological property under consideration in this situation is not that arising from a single, isolated structure but rather one from a drug-receptor pair. A valid comparison would study the mathematical properties of the drug-receptor pairs for both antipodal structures. This sort of comparison would reveal that the two systems were not alike at all and were in fact quite different.

OUTLINE OF GRAPH-THEORETICAL SCHEMES

In order to proceed with our treatment, it will be necessary to accomplish three tasks, viz (i) to represent all the structures of interest in mathematical terms, i.e. as chemical graphs; (ii) to prescribe some comparability test that will indicate how much two given structures differ; and (iii) to recognize the significant components in the molecules considered. The first two of these tasks have been addressed in many publications discussing the graph-theoretical approach to structure-activity studies (Randić and Wilkins, 1979a, 1979b; Wilkins and Randić, 1980; Wilkins et al., 1981; Randić, 1985a; Jerman-Blazić et al., 1985), and thus need not be further elaborated here. For our purposes we shall represent chemical compounds by their molecular graphs with the nonessential hydrogen atoms suppressed in the customary fashion (Trinajstić, 1983), though we shall also discuss the representation of structures by appropriately weighted path numbers. Our main focus of attention, however, will be on task (iii) and the means of identifying the essential components within a set of structurally related compounds displaying differing pharmacological activities.

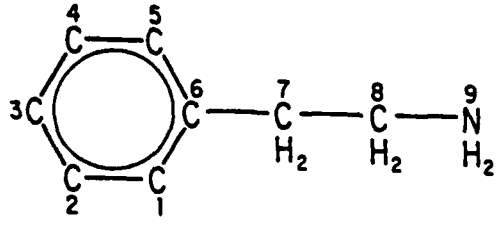


FIG. 1. Graph of the molecule of phenylethylamine.

Let us first consider the compound shown in Figure 1 and assign an arbitrary numbering to the atoms therein.

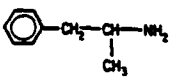
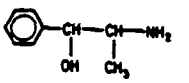
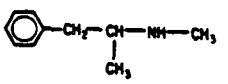
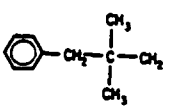
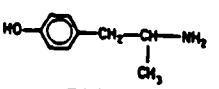
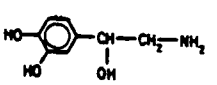
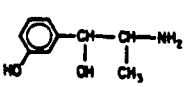
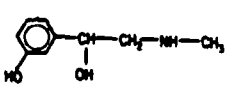
Drug		Name	Pharmacological Classification	Therapeutic Application
1.		Amphetamine	CNS Stimulant	Antidepressant Anorexiant Narcolepsy Minimal Brain Dysfunction
2.		Phenylpropanoamine	α-Agonist CNS Stimulant	Decongestant Anorexiant
3.		Methamphetamine	CNS Stimulant	Anorexiant Antidepressant Narcolepsy Minimal Brain Dysfunction
4.		Phentermine	CNS Stimulant	Anorexiant
5.		Hydroxyamphetamine	α-Agonist	Antihypotensive
6.		Levarterenol	α-Agonist	Antihypotensive Vasoconstrictor
7.		Metaraminol	α-Agonist	Antihypotensive
8.		Methamphetamine	CNS Stimulant	Anorexiant Minimal Brain Dysfunction Antidepressant Narcolepsy

FIG. 2. Set of graphs of molecules closely related to the phenylethylamine molecule.

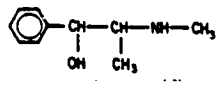
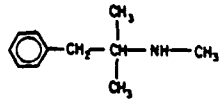
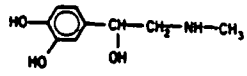
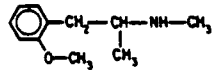
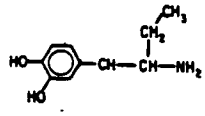
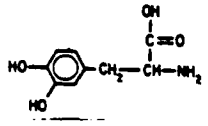
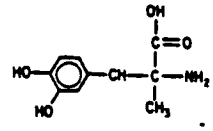
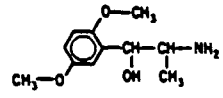
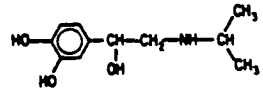
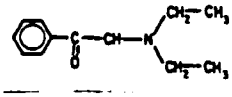
9.		Ephedrine	α-Agonist CNS Stimulant	Antihypotensive Decongestant Bronchodilator Antiarrhythmic
10.		Mephentermine	CNS Stimulant α-Agonist	Antihypotensive
11.		Epinephrine	α-Agonist β-Agonist	Antihypotensive Bronchodilator Decongestant Antiarrhythmic
12.		Methoxyphenamine	β-Agonist	Bronchodilator
13.		Ethylnorepinephrine	β-Agonist	Bronchodilator
14.		Levodopa	CNS Agent	Antiparkinsonian
15.		Methyldopa	CNS Agent	Antihypertensive
16.		Methoxamine	α-Agonist	Antiarrhythmic Antihypotensive
17.		Isoproterenol	β-Agonist	Antiarrhythmic Bronchodilator
18.		Diethylpropion	CNS Stimulant	Anorexiant

FIG. 2. Set of graphs of molecules closely related to the phenylethylamine molecule. (continued)

This particular compound is closely related to all of the compounds depicted in Figure 2. The simplified molecular graphs we are using here do not differentiate between aromatic and aliphatic C-C bonds, and neither do they distinguish between single C-C and C-N bonds. Our ultimate interest concerns the set of compounds shown in Figure 2, all of which possess certain identical structural features: a phenyl ring and a nitrogen atom removed from the ring by three bonds. In the present context, differentiation between bond types turns out to be of no great consequence; in fact it was recently demonstrated (Grossman et al., 1985) that characterization of structures by means of weighted paths (where heteroatoms were given differing weights) was rather insensitive to the actual choice of weights.

Our characterization of the compounds illustrated above will be based on the counts of paths of different lengths. A path of length k will represent a fragment containing k consecutive bonds, i.e. a chain of length k. By convention, paths of length zero represent atoms, paths of length one count the number of bonds, paths of length two count the number of pairs of consecutive bonds, and so on. In Table 1 we present the counts for paths of increasing length for each of the atoms in the compound depicted in Figure 1. These counts were arrived at by making use of the ALLPATH program (Randić et al., 1979). For compounds having no more than a single ring and a few atoms, carrying out the counts is not particularly onerous, yet, even for bicyclic systems, it becomes impractical to perform the counts by hand. The last row in Table 1, having the entries 9 9 10 11 12 12 6 4 2, gives the path counts for the molecule as a whole. These counts can be readily derived from the data on the individual atoms if it is remembered that all the paths (except those of zero length) have been counted twice -- once for each end atom. The molecule in Figure 1 thus has 9 atoms, 9 bonds, 10 adjacent pair bonds, 11 sets of three consecutive bonds, and so on.

The advantage of using path numbers, as opposed to molecular fragments, such as bonds or small atomic groups, is that the path numbers retain some information on the nonlocal connectivity within the structure. However, it is evident that the number of paths of intermediate length dominate in the path counts, a fact which may result in similarities among compounds due to local characteristics being obscured. From studies on isomeric variations in the physicochemical properties of species (Randić and Wilkins, 1979c, 1979d; Randić and Wilkins, 1980; Randić and Trinajstić, 1982), it has been established that shorter paths, especially those of lengths two and three, play a crucial role. It would thus appear desirable to introduce weighting of the paths in order that the dominant role of the more abundant paths of intermediate length can be counteracted, and the role of the shorter paths given greater prominence. As weighting of paths based on a differentiation of bond types has been found effective for such purposes (Menon and Cammarata, 1977, Randić, 1984a; Randić, 1985a), we shall adopt this approach here.

TABLE 1. The output of the ALLPATH program for the molecule in Fig. 1.

0	1	0	0	0	1	0	0	0
1	0	1	0	0	0	0	0	0
0	1	0	1	0	0	0	0	0
0	0	1	0	1	0	0	0	0
0	0	0	1	0	1	0	0	0
1	0	0	0	1	0	1	0	0
0	0	0	0	0	1	0	1	0
0	0	0	0	0	0	1	0	1
0	0	0	0	0	0	0	1	0
---	---	---	---	---	---	---	---	---
1								
1	2	3	3	3	2	1	1	1
---	---	---	---	---	---	---	---	---
2								
1	2	2	3	3	4	1	1	0
---	---	---	---	---	---	---	---	---
3								
1	2	2	2	4	4	2	0	0
---	---	---	---	---	---	---	---	---
4								
1	2	2	3	3	4	1	1	0
---	---	---	---	---	---	---	---	---
5								
1	2	3	3	3	2	1	1	1
---	---	---	---	---	---	---	---	---
6								
1	3	3	3	2	2	0	0	0
---	---	---	---	---	---	---	---	---
7								
1	2	3	2	2	2	2	0	0
---	---	---	---	---	---	---	---	---
8								
1	2	1	2	2	2	2	2	0
---	---	---	---	---	---	---	---	---
9								
1	1	1	1	2	2	2	2	2
---	---	---	---	---	---	---	---	---
9	9	10	11	12	12	6	4	2

TOTAL NUMBER OF PATHS: 75

To carry out the weighting, each bond is classified as being of (m,n) type, where m and n are the numbers of edges emanating from each of the terminal vertices of the bond in question. For all bond types (m,n), a weight of $(m \times n)^{-\frac{1}{2}}$ is assigned to each bond, following the same procedure adopted in computing the connectivity indices of molecules (Randić, 1975). This weighting procedure many be used in conjunction with the widely available ALLPATH program (Randić et al., 1980) provided the weights are entered as input. Alternatively, a subroutine may be added to the existing program to automatically introduce weightings in the counting process (Randić, 1985b). In Table 2 are listed the counts for the weighted paths of the compound depicted in Figure 1. The results represent the printed output of a modified ALLPATH program.

In addition to the path numbers, i.e. the numbers of paths of different length, the output also gives for each atom the total of all the paths pertaining to that atom. Thus, for atom 1 this total is 3.011 whereas for atom 2 it is only 2.989, and so on. It appears that these 'atomic' numbers are able to differentiate between atomic environments; they may therefore be referred to as atomic identification (ID) numbers. The last line of the output, reproduced here with the numbers truncated to three decimal places:

9 4.431 2.215 1.098 0.574 0.276 0.071 0.025 0.007

represents the (weighted) path counts for the molecule as a whole. The total number of paths, 17.7005, has been termed the molecular ID (identification number), and has been shown to be a highly discriminating (Randić, 1984a) though not unique (Szymanski et al., 1985) index.

COMPARISON OF DIFFERENT STRUCTURES

For the molecule illustrated in Figure 1, Table 2 provides a set of graph invariants that may be used in comparing similar data on a variety of other structures. A sequence, such as the above list of paths of different lengths, or a set of numbers, such as the list of all 'atomic' numbers, clearly offers a broader basis for the comparison of structures than (say) a single topological index, such as the connectivity index originally introduced to discuss the branching in alkane molecules and variations among the physicochemical properties of isomeric species (Randić, 1975). As will be evident, even the use of a single number as a descriptor, e.g. a partial sum of selected 'atomic' numbers, can yield extremely useful information from the comparisons between structures. To those not especially well versed in chemical graph theory, it may come as something of a surprise that a single number is able to capture so much of the essential structural information associated with chemical species, though this observation has been amply corroborated in the manifold applications of the connectivity index. Such numbers, which to the uninitiated may appear to be ad hoc in origin, are in fact based upon well-defined and important structural invariants.

The information presented in Table 2 can be used in several different ways. The path numbers for each compound may be viewed as the components of a vector and the degree of similarity existing between different vectors then established. The similarity can be defined in terms of the Euclidean distance between the position vectors in n-dimensional space. This general type of analysis has already been applied to the dopamines, benzomorphans, barbiturates, and aminotetralins (Randić and Wilkins, 1979c; Randić and Wilkins, 1979d; Randić and Wilkins, 1980; Randić and Trinajstić, 1982). Alternatively, structures may be represented by sets comprised of the relevant atomic path sums, where the summation is restricted to selected atoms only, as illustrated in the search for optimal antitumor drugs (Randić, 1985a). Use of the molecular ID numbers for the purpose of clustering compounds together can be made only on the basis of similarities existing among the individual ID values. In the cases of several therapeutically valuable antihistamines, anticholinergics, antipsychotics, antidepressants, analgesics, and antiparkinsonians, however, surprisingly good classifications based solely on this single structural parameter have been obtained (Randić, 1984b).

As will be evident from Table 2, ID numbers are size-dependent. For the compounds we have considered here, each atom contributes around 2.25 to 3.00 to the ID number. It seems quite likely that such 'size' effects may obscure some of the finer structural differences existing among the compounds illustrated in Figure 2. The effect may be especially pronoucned here because all the molecules concerned are relatively small, i.e. they contain no more than 10-15 atoms each, not counting the suppressed hydrogen atoms. In the following section, we shall select a fragment present in all the compounds of Figure 2. Comparison of the compounds will be based solely upon the characteristics of the atoms common to all the structures considered.

CLUSTERING OF THE THERAPEUTICALLY RELATED SPECIES

The compounds represented in Figure 2, all of which are therapeutically very efficacious, form a subset

TABLE 2. The output of the ALLPATH program with weighting of bonds for the molecule in Fig. 1

```
- - - - -
1
1                    .908248291          .583333333          .291666667          .163092232
.0463488515          .0104166667         5.20833334E-03      3.68284782E-03
3.01199722
- - - - -
2
1                    1                   .454124145          .291666667          .13436437
.0919627826          .0104166667         7.36569564E-03      0
2.98990033
- - - - -
3
1                    1                   .5                  .204124145          .166666667
.0833333334          .0294627826         0                   0
2.98358693
- - - - -
4
1                    1                   .454124145          .291666667          .13436437
.0919627826          .0104166667         7.36569564E-03      0
2.98990033
- - - - -
5
1                    .908248291          .583333333          .291666667          .163092232
.0463488515          .0104166667         5.20833334E-03      3.68284782E-03
3.01199722
- - - - -
6
1                    1.22474487          .612372436          .348461713          .102062073
.0510310363          0                   0                   0
3.33867213
- - - - -
7
1                    .908248291          .686886724          .166666667          .0833333334
.0416666667          .0208333333         0                   0
2.90763502
- - - - -
8
1                    1.20710678          .204124145          .1666666667         .0833333334
.0416666667          .02083333333        .0104166667         0
2.73414759
- - - - -
9
1                    .707106781          .353553391          .144337567          .11785113
.0589255651          .0294627826         .0147313913         7.36569564E-03
2.43333431
- - - - - - - - - - -
9                    4.43185765          2.21592583          1.09846171          .57407987
.276623268           .0711294493         .025148058          7.36569564E-03


TOTAL NUMBER OF PATHS: 17.7005855
```

TABLE 3. The atomic ID number for the nine atoms common to all 18 structures of Fig. 2.

Drug	Atom positions: 1	2	3	4	5	6	7	8	9
1	3.016	2.992	2.985	2.992	3.016	3.349	2.933	2.993	2.394
2	3.015	2.992	2.985	2.992	3.015	3.347	3.156	2.967	2.379
3	3.025	2.997	2.989	2.997	3.025	3.369	2.982	3.112	2.693
4	3.017	2.993	2.9986	2.993	3.017	3.351	2.937	3.226	2.363
5	3.016	2.980	3.198	2.980	3.016	3.357	2.936	2.994	2.395
6	3.001	3.192	3.185	2.977	3.013	3.349	3.138	2.703	2.411
7	3.004	3.203	2.971	2.990	3.018	3.352	3.158	2.967	2.380
8	3.010	3.206	2.973	2.993	3.024	3.365	3.196	2.849	2.705
9	3.201	2.995	2.988	2.995	3.021	3.360	3.196	3.086	2.682
10	3.023	2.996	2.989	2.996	3.023	3.366	2.974	3.329	2.670
11	3.009	3.196	3.188	2.981	3.022	3.369	3.197	2.850	2.705
12	3.350	3.028	3.013	3.019	3.049	3.408	2.998	3.119	2.696
13	3.009	3.196	3.188	2.981	3.022	3.370	3.199	3.088	2.449
14	3.009	3.196	3.188	2.981	3.021	3.369	2.962	3.054	2.430
15	3.008	3.196	3.188	2.981	3.021	3.368	2.960	3.279	2.389
16	3.362	3.045	3.040	3.341	3.074	3.410	3.177	2.974	2.383
17	3.014	3.198	3.190	2.984	3.027	3.381	3.233	2.936	2.877
18	2.983	2.973	2.970	2.973	2.983	3.272	3.138	3.083	3.155

of a collection of compounds investigated by Menon and Cammarata (1977) using pattern recognition techniques. From their collection of almost 40 compounds, we have selected 18 compounds whose molecules contain no cycles other than a single phenyl group, no chlorine atoms as substituents, and no quarternary nitrogen atoms. All of the selected compounds are closely related structurally: apart from having a phenyl ring, they all have a nitrogen atom three bonds removed from this ring. They do differ, however, in the number, type, and position of the various substituents they contain, namely the hydroxyl group, the methyl or ethyl groups, and occasionally the carbonyl group. In Table 3 a partial path characterization of these compounds is presented, with only the atomic path numbers appearing for the nine atoms common to all of the compounds. Inspection of Figure 2 reveals that the structure we have depicted in Figure 1 is the largest fragment common to all the 18 compounds.

The partial sums of the atomic path numbers for the nine common atoms are reported in Table 4. The nine atoms have now been partitioned into two groups: the six atoms constituting the phenyl ring are considered separately (for reasons which will become apparent later). The remaining entries in Table 4 are for the three atoms forming the side chain (including the nitrogen); the totals for the nine-atom fragment are also given in each case. These latter totals we shall refer to as the fragment ID numbers. Analysis of the fragmentID numbers indicates that the 'size' effect mentioned above has now been eliminated. Use of the fragment IDs to order the compounds, however, leads to the disappointing result that such ordering produces no significant pharmacological classification of the compounds.

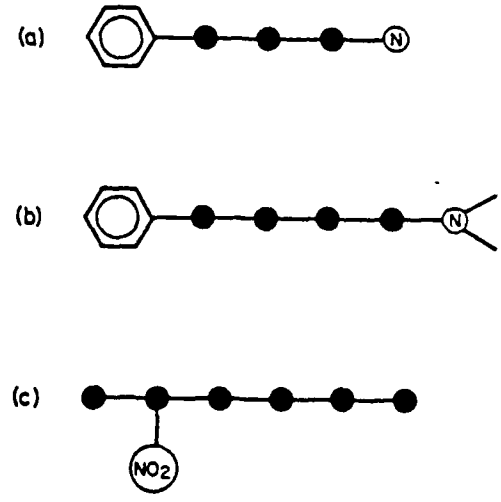


FIG. 3. The fragments thought to be essential for the activity of (a) the morphins, (b) neuroleptics, (c) mutagenic nitroarenes.

The nine-atom fragment is comparable in size with a number of other groupings identified as performing an essential pharmacophoric role in various drug

TABLE 4. Partial sums of atomic path numbers for the ring and other atoms in the fragment (see Fig. 1).

Drug	Ring ID	Side Chain ID	Fragment ID
1 Amphetamine	18.353	8.321	26.675
2 Phenylpropanolamine	18.347	8.503	26.850
3 Metamphetamine	18.404	8.788	27.193
4 Phentermine	18.357	8.527	26.880
5 Hydroxyamphetamine	18.548	8.326	26.875
6 Levarterenol	18.719	8.253	26.973
7 Metaraminol	18.541	8.506	27.047
8 Phenylephrine	18.573	8.751	27.324
9 Ephedrine	18.382	8.965	27.347
10 Mephentermine	18.396	8.975	27.371
11 Epinephrine	18.767	8.753	27.521
12 Methoxyphenamine	18.870	8.814	27.685
13 Ethylnorepinephrine	18.768	8.736	27.505
14 Levodopa	18.766	8.446	27.213
15 Methyldopa	18.764	8.629	27.393
16 Metoxamine	19.274	8.535	27.809
17 Isoproterenol	18.769	9.047	27.843
18 Diethylpropion	18.155	9.376	27.532

molecules. For instance, the empirical 'morphine rule' fragment (Lednicer and Mitscher, 1977), the fundamental structure proposed for neuroleptic action (Janssen, 1964), and the significant fragment in the mutagenic nitroarenes (Klopman and Rosenkranz, 1984) are all of a similar size and each is claimed to be specific. The three fragments are illustrated in Figure 3. In our case, the nine-atom group we consider is clearly pharmacologically active, though its action is nonspecific. What is required at this point is a finer differentiation among the 18 compounds under consideration.

In Figure 4 a histogram is presented based only upon the ring ID values, i.e. the values of the atomic path sums for the six atoms constituting the phenly ring in the 18 compounds of interest. Rather surprisingly, there is now a very evident clustering of all the central nervous system (CNS) simulants (which have lower values of the ring ID), and similar clusterings for the β-agonists and the CNS agents. The latter group,which clusters in the central region of Figure 4, has too few compounds within it to give any great statistical significance to this particular finding. Moreover, by contrast, there is a wide scatter for the α-agonists over the whole range of ring ID values.

These observations, which are highly interesting, might have escaped attention altogether if only a visual inspection of the structures had been made. Clustering

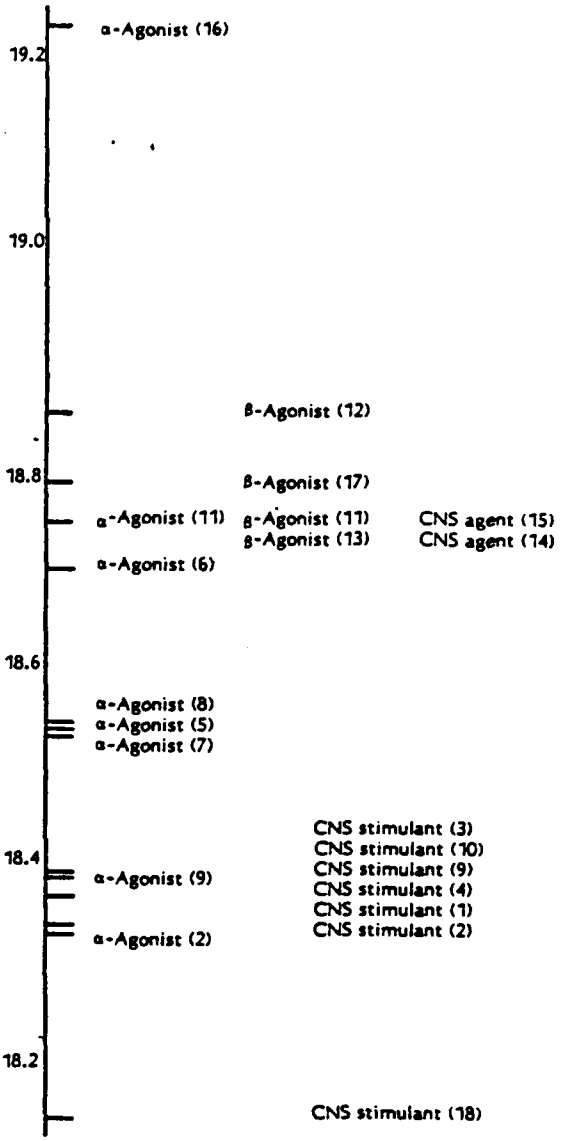


FIG. 4. Classification of the bioactivity of the 18 compounds considered based on their ring ID numbers.

of all the CNS stimulants within the range of ring ID values lying between 18.35 - 18.40, which represents a small interval compared to the full range of possible ring ID values (from around 18.00 to 19.25) for the compounds under study, indicates that rings lying within this narrow range contain a specific structural component essential for that particular type of pharmacologic activity. This interval relates, of course, only to unsubstituted phenyl rings, and careful inspection of the molecular diagrams might have revealed that such rings are essential for CNS stimulants. In the case of the β-agonists, a phenyl ring with two substituent hydroxyl groups appears to be essential, and the

same seems to apply in the CNS agents. These can certainly be interpreted as positive results, although the scatter of the α-agonist ring ID values over the whole range of possible values must be seen as a negative result. The finding that for α-agonistic-type activity substitution (by hydroxyl groups) of the phenyl ring may occur is without special significance.

CONCLUDING REMARKS

In this presentation only one particular aspect of the graph-theoretical approach to quantitative structure-activity relationships has been examined. After visually identifying a common nine-atom fragment among a group of therapeutically valuable drugs, attention was focused on one critical component of the fragment. This component was a ring which played a major role in discriminating between structures for pharmacological classification purposes. Thus, it is important to recognize that not only fragments may be responsible for pharmacological action, but that such fragments may need to be further subdivided in order to obtain a good correlation between a given structure and its function. Even a negative result, such as the discovery that the behavior of a fragment is insensitive to selective substitution; is of considerable interest in drug design studies. For one thing it suggests that the least expensive derivative may be used for any substitution which is irrelevant, provided that other factors, such as toxicity and dosage, remain unchanged.

If the compounds listed in Figure 2 are regarded as lead compounds, the analysis presented here can serve to indicate both productive and unfruitful approaches to the design of enhanced drugs. In the case of CNS stimulants, for instance, it is clear that it would be undesirable to attempt to substitute the phenyl ring, whereas for α-agonists this would be an allowed possibility. The actual direction adopted will, of course, depend very heavily on which particular standards are recommended as optimal. Compounds which appear most promising would in general differ least in the essential fragment, that is to say the mathematical characterization of the fragments should differ least from that of the lead molecule. Although it is not unreasonable to adopt the approach pursued in several similar studies reported previously (Menon and Cammarata, 1977; Trinajstić, 1983; Randić, 1985a), one has now gained important additional insights. It is much better known which part of the overall molecular characterization is most crucial. By clustering together structures that are most similar in their more significant structural details, some uncertainties in the search for optimal drugs can certainly be eliminated.

ACKNOWLEDGMENTS

Ames Laboratory - DOE is operated for the U.S. Department of Energy by Iowa State University under contract W-7405-ENG-82 and M.R. was supported in part by the U.S. Office of the Director. D.H.R. thanks the U.S. Office of Naval Research for partial support of this project.

REFERENCES

Bonchev, D. (1983). Information-Theoretic Indices for Characterization of Chemical Structures. Research Studies Press, Chichester, United Kingdom.

Burger, A. (1983). A Guide to the Chemical Basis of Drug Design. Wiley, New York.

Crum Brown, A. and T.R. Fraser (1868). On the connection between chemical constitution and physiological action. I and II. Trans. Roy Soc. Edinburgh, 25, 151-302, 693-739.

Ehrlich, P. (1885). Das Sauerstoff-Bedurfniss des Organismus: eine farbenanalytische Studie. Hirschwald, Berlin.

Fischer, E. (1984). Einfluss der Configuration auf die Wirkung der Enzyme. Chem. Ber. 27, 2985-2993.

Grossman, S.C., B. Jerman-Blazić, and M. Randić (1985). A graph-theoretical approach to quantitative structure-activity relationship. Int. J. Quant. Chem., Quant. Biol. Symp., In press.

Hansch, C. (1969). A quantitative approach to biochemical structure-activity relationships. Accts. Chem. Res., 2, 232-239.

Janssen, P.A. (1964). As quoted by Kaufman, J.J. and E. Kerman (1974). The structure of psychotropic drugs. Int. J. Quant. Chem., Quant. Biol. Symp. 1, 259-287. Int. J. Quant. Chem., Quant. Biol. Symp. 1, 259-26.

Jerman-Blazić, B., M. Randić, and J. Zerovnik (1985). A new approach in computer-aided derivation of QSAR. In J.K. Seydel (Ed.), QSAR and Strategies in the Design of Bioactive Compounds, VCH Press, Weinheim, Germany, pp. 39-49.

Jerman-Blazić, B. and M. Randić (1983). Modelling molecular structures for computer-assisted studies of drug structure-activity relations. In Modelling and Simulation, AMSE Press, Tassin, France, 5, 161-174.

Kier, L.B. and L.H. Hall (1976). Molecular Connectivity in Chemistry and Drug Research. Academic Press, New York.

Klopman, G. and H.S. Rosenkranz (1984). Structural requirements for the mutagenicity of environmental nitroarenes. Mutation Res., 126, 227-238.

Langmuir, I. (1919). Isomorphism, isosterism and covalence. J. Am. Chem. Soc. 41, 1543-1 559.

Lednicer, D. and L.A. Mitscher (1977). The Organic Chemistry of Drug Synthesis, Vol. 1. Wiley-Interscience, New York, pp. 286-293.

Menon, G.K. and A. Cammarata (1977). Pattern recognition II: Investigation of structure-activity relationships. J. Pharm. Sci., 66, 304-314.

Randić, M. (1975). On characterization of molecular branching. J. Am. Chem. Soc., 97, 6609-6615.

Randić, M. (1984a). On molecular identification numbers. J. Chem. Inf. Comput. Sci., 24, 164-175.

Randić, M. (1984b). Nonempirical approach to structure-activity studies. Int. J. Quant. Chem., Quant. Biol. Symp., 11, 137-153.

Randić, M. (1985a). Graph-theoretical approach to structure-activity studies: search for optimal antitumor compounds. In R. Rein (Ed.), Molecular Basis of Cancer, Part A, Liss Publishers, pp. 309-318.

Randić, M. (1985b). Program written in BASIC for the Apple IIe personal computer. Available upon request, provided no commercial use is intended.

Randić, M., G.M. Brissey, R.B. Spencer, and C.L. Wilkins (1979). Search for all self-avoiding paths for molecular graphs. Computers & Chem., 3, 5-13.

Randić, M. G.M. Brissey, R.B. Spencer, and C.L. Wilkins (1980). Use of self-avoiding paths for characterization of molecular graphs with multiple bonds. Computers & Chem., 4, 27-43.

Randić, M. G.A. Krans, and B. Jerman-Blazić (1983). Ordering of graphs as an approach to structure-activity studies. Studies Phys. Theor. Chem., 28, 192-205.

Randić M. and C.L. Wilkins (1979d). Graph-theoretical approach to recognition of structural similarity in molecules. J. Chem. Inf. Comput. Sci., 19, 31-37.

Randić, M. and C.L. Wilkins (1979b). Graph-theoretical study of structural similarity in benzomorphans. Int. J. Quant. Chem., Quant. Biol. Symp., 6, 55-71.

Randić, M. and C.L. Wilkins (1979c). On a graph-theoretical basis for ordering of structures. Chem. Phys. Lett., 63, 332-336.

Randić, M. and C.L. Wilkins (1979a). Graph-theoretical ordering of structures as a basis for systematic searches for regularities in molecular data. J. Phys. Chem., 83, 1525-1540.

Randić, M. and C.L. Wilkins (1980). Graph-theoretical analysis of molecular properties. Isomeric variations in nonanes. Int. J. Quant. Chem. 18, 1005-1027 (1980).

Randić, M. and N. Trinajstić (1982). On isomeric variations in decanes. Math. Chem., 13, 271-290.

Rouvray, D.H. (1973). The search for useful topological indices in chemistry. Amer. Sci., 61, 729-735.

Stuper, A., W. Brugger, and J. Jurs (1979). Computer-Assisted Studies of Chemical Structure. Wiley, New York.

Syzmanski, K. W.R. Muller, J.V. Knop, and N. Trinajstić (1985). Private communication.

Thornber, C.W. (1979). Isosterism and molecular modification in drug design. Chem. Soc. Revs. 8, 563-580.

Trinajstić, N. (1983). Chemical Graph Theory, Vols I and II. CRS Press, Boca Raton, Florida.

White, H.J. and S. Tauber (1969). Systems Analysis. Saunders, Philadelphia.

Wilkins, C.L. and M. Randić (1980). A graph-theoretical approach to structure-property and structure-activity correlations. Theor. Chim. Acta, 58, 45-68.

Wilkins, C.L., M. Randić, S.M. Schuster, R.S. Markin, S. Steiner, and L. Dorgan (1981). A graph-theoretical approach to quantitative structure-activity/reactivity studies. Anal. Chim. Acta, 133, 637-645.

END

DTIC

8-86